

6

GENOM CZŁOWIEKA

Alina Woźniak, Celestyna Mila-Kierzenkowska

- **genom – budowa i funkcja**
- **genom mitochondrialny**
- **projekt poznania genomu ludzkiego (HGP)**
- **mapowanie genomu**
- **pytania**
- **piśmiennictwo**

Termin „genom” został po raz pierwszy wprowadzony w 1920 r., przez niemieckiego botanika Hansa Winklera w celu określenia haploidalnego zespołu chromosomów wyodrębnionego z *Eukaryota*. Obecnie pojęcie genomu określa sumę wszystkich kodujących i niekodujących sekwencji DNA zawartych w haploidalnej liczbie chromosomów.

W komórkach *Eukaryota* występuje również pozajądrowy materiał genetyczny w postaci mitochondrialnego (mtDNA) i plastydowego DNA, które stanowią odrębne genomy. Genomem określa się też materiał genetyczny bakterii oraz wirusów, bez względu na to, czy wirus zawiera DNA czy RNA.

Badaniem genomów zajmuje się „genomika” – termin wprowadzony w 1987 r. przez V.A. McKusicka i F.H. Ruddle’a.

6.1

Rodzaje i wielkość genomu

U *Prokaryota* oraz u niektórych organizmów eukariotycznych (np. u trutni pszczoł) w komórce występuje pojedynczy genom, gdyż są one haplontami. U większości *Eukaryota* w toku ewolucji doszło do wykształcenia komórek o podwójnym genomie (diploidalnych).

Wyróżnia się dwa typy genomów: RNA i DNA. U *Prokaryota* i *Eukaryota* występują genomy typu DNA, natomiast u wirusów – genomy typu DNA lub RNA. Jednym z najmniejszych genomów jest genom poliovirusa.

Składa się on z około 7400 par zasad nici (+)RNA.

DNA genomu jądrowego człowieka liczy około 3000 Mb (3 miliardy par zasad), ma około 100 cm długości i masę cząsteczkową 3×10^9 kD. U człowieka najmniejszy chromosom – pary 21 zawiera 47 Mb, a największy – pary 1 zawiera 247 Mb. Dla porównania DNA *E. coli* zawiera 4,6 Mb (1,35 mm długości), *Drosophila melanogaster* 140 Mb (56 mm długości), a genom ryby dwudysznej – 102 000 Mb (34,7 metra długości). Najprostsze *Eukaryota*, takie jak grzyby, mają najmniejsze genomy, a kręgowce i rośliny nasienne jedne z największych. Okazuje się jednak, że stopień rozwoju ewolucyjnego danego organizmu nie koreluje z ilością DNA w jego

komórkach. Genom salamandry (płazy) jest np. 100 razy większy niż genom ptaków, a 10 razy większy niż genom ssaków. Z kolei wielkość genomu ssaków odpowiada w przybliżeniu wielkości genomu pierwotniaków.

6.2

Organizacja genomu – izochory

Genom kręgowców, w tym również genom człowieka, jest mozaiką izochor. Izochory to długie, liczące ponad 300 kb (1 kb = 1000 par zasad) regiony DNA o homogenym (jednolitym) składzie zasad azotowych. Izochory podzielone są na kilka rodzin. Poszczególne rodziny różnią się sumą i zawartością par zasad G-C i C-G (G+C) oraz gęstością wykazaną podczas wirowania DNA w stężeniu Cs_2SO_4 w obecności soli srebra, pełniącej rolę specyficznego liganda dla sekwencji bogatych w zasady G+C. Około 62% genomu człowieka reprezentują rodziny izochor „lekkich”, ubogie w G+C (*L1* i *L2*).

Rodziny izochor „ciężkich”, bogate w zasady G+C (*H1* i *H2*) i bardzo bogate w G+C (*H3*), stanowią odpowiednio 31% i 3–4% genomu. Pozostałe 3–4% genomu składa się z satelitarnego i rybosomowego DNA (rDNA), który nie wchodzi w skład izochor, ale wygląda jak izochora z powodu jednolitego składu zasad.

Obecność izochor w genomowym DNA stwierdzono po raz pierwszy w latach siedemdziesiątych ubiegłego stulecia, podczas wirowania DNA w określonym stężeniu Cs_2SO_4 . Niestety, nie można dokonać rozdzielania DNA na izochory typowymi metodami izolacji genomowego DNA. Tymi metodami uzyskuje się tylko fragmenty izochor długości 50–100 kb. Nie udało się również do tej pory zsekwencjonowanie fragmentu DNA będącego chociaż jedną całą izochorą.

Geny w ludzkim genomie nie są rozmieszczone losowo. Wykazano, że w większości przypadków badane geny były obecne w rzadko występujących, bogatych w G+C

izochorach. W izochorach ubogich w zasady G+C gęstość genów jest niska i zwiększa się ze wzrostem zawartości G+C w rodzinach izochor *H1* i *H2*. Najwyższa gęstość genów występuje w rodzinie izochor *H3*, która wykazuje ponad 20-krotnie większą gęstość genów od izochor ubogich w G+C. Rodzina izochor *H3* została w związku z tym nazwana „rdzeniem genomu”. Zawartość genów w izochorach *L1* i *L2* wynosi 4%, a pozostałe 96% to sekwencje niekodujące, natomiast w izochorach *H1* i *H2* – 20% to sekwencje kodujące. Najwięcej genów, aż 76%, występuje w najcięższych izochorach *H3*. Zawartość zasad G+C w genomie jest więc miarą gęstości genów.

Wzrostowi zawartości genów w izochorach wraz ze wzrostem zawartości zasad G+C towarzyszy również zwiększona obecność wysp CpG. Wyspy te są regionami niemetylowanego DNA w jednej nici wielkości ok. 1 kb, w których bardzo często występuje deoksycytidyna i deoksyguanozyna, tworzące dinukleotydy CpG. Wyspy CpG nie zawierają żadnej dłuższej specyficznej sekwencji nukleotydów, która umożliwiałaby ich identyfikację. Są więc umownie określane jako fragmenty DNA silnie trawione restryktazami specyficznymi dla sekwencji CpG albo jako regiony o wysokiej zawartości zasad G+C (ok. 65%). Wyspy CpG to prawdopodobnie miejsca o charakterze promotorów, z którymi oddziałują czynniki transkrypcyjne. Znajdują się one głównie w cytotogenetycznych prążkach R, które obejmują 50% genomu. Przybliżona liczba wysp CpG w haploidalnym genomie wynosi 45 000.

Istnieją różnice w zawartości poszczególnych rodzin izochor w prążkach chromosomalnych. Wykazano, że:

- prążki G są utworzone wyłącznie z ubogich w zasady G+C izochor, z małym udziałem izochor bogatych w G+C z rodziny *H1*, są więc dość jednorodne pod względem składu, gdyż izochory lekkie nieznacznie się różnią pod względem zawartości zasad G+C;
- prążki R zawierają izochory obejmujące fragmenty DNA o różnej liczbie zasad

G+C, są więc bardziej zróżnicowane; do prążków R zalicza się:

- prążki T zbudowane przeważnie z rodzin izochor *H2* i *H3*,
- prążki R' składające się z prawie jednakowej liczby izochor bogatych w G+C (głównie rodziny *H1*), i ubogich w G+C. Prążki R' można podzielić na dwie podklasy zwane prążkami T' i R". Prążki T' zawierają głównie izochory *H3* i *H2*, zaś prążki R" ich nie mają. Prążki R zawierają zatem trzy rodzaje prążków: T, T' i R".

W prążkach T znajduje się około 46–58% wszystkich genów człowieka. Stanowią one jednocześnie tę część genomu, która wykazuje najwyższą aktywność transkrypcyjną i rekombinacyjną. Prążki T są również regionem najczęstszego występowania wysp CpG.

Poszczególne rodziny izochor są związane ze swoistą strukturą chromatyny. Struktura chromatyny jest „otwarta” w izochorach bogatych w G+C, czyli w rodzinach *H2* i *H3*. „Otwarta” struktura chromatyny charakteryzuje się łatwym dostępem deoksyrybonukleaz (DNA-z) spowodowanym niedoborem lub brakiem histonu H1, acetylacją histonu H3 i H4 oraz większym odstępem między nukleosomami.

Różnice w budowie genomu kręgowców dotyczą także organizacji izochor. Genom kręgowców zmiennocieplnych charakteryzuje mała różnorodność kompozycyjna, ich izochory bogate w zasady G+C występują rzadziej niż bogate w zasady G+C izochory kręgowców stałocieplnych.

Podczas ewolucyjnego przejścia od gadów do ptaków i ssaków (dwie odrębne linie rozwojowe) nastąpił wzrost zawartości zasad G+C, zwłaszcza w izochorach z rodzin *H1–H3*. Zmianom tym towarzyszyło pojawienie się prążków R widocznych po wybarwieniu chromosomów. Następowaly także zmiany kompozycyjne w sekwencji kodującej, zwłaszcza w trzeciej pozycji kodonu. Porównując geny homologiczne człowieka i *Xenopus*, wykazano, że u człowieka geny mają wyższą zawartość zasad G lub C w trzeciej pozycji kodonu.

6.3

Pierwszorzędowa struktura genomu człowieka

Występowanie w genomie określonych sekwencji zasad decyduje o strukturze pierwszorzędowej genomu. Jest ona odmienna u *Prokaryota* i *Eukaryota*. U spokrewnionych gatunków również obserwuje się różnice w organizacji sekwencji zasad w genomie. Około 30% genomu jądrowego człowieka stanowią sekwencje ulegające transkrypcji i sekwencje związane z genami (eksony, introny, pseudogeny, fragmenty genów, sekwencje regulatorowe, sekwencje początkowe i końcowe genów). Pozostałą część stanowi pozagenowy DNA obejmujący sekwencje unikatowe oraz sekwencje powtórzone (repetytywne).

6.3.1

Geny i sekwencje związane z genami

Geny i sekwencje funkcjonalnie związane z genami obejmują u człowieka 900 Mb, w tym DNA kodujący zawiera 90 Mb, a DNA niekodujący, obejmujący m.in. introny i pseudogeny, zawiera pozostałe 810 Mb. Geny *Eukaryota* mają strukturę nieciągłą, co oznacza, że informacja o sekwencji aminokwasów nie jest ciągła, lecz przerywana odcinkami niekodującymi. Sekwencje kodujące nazywane są eksonami, niekodujące zaś intronami. W większości przypadków introny są dłuższe niż eksony.

6.3.2

Pozagenowy DNA

Pozagenowy DNA obejmuje u człowieka 2100 Mb, w tym sekwencje powtórzone stanowią 420 Mb, a unikatowe 1680 Mb.

6.3.2.1

Unikatowy DNA

Sekwencje unikatowe w genomie człowieka stanowią ok. 80% pozagenowego DNA. Występują w genomie głównie w postaci jednej kopii.

6.3.2.2

DNA powtórzone

W genomie *Eukaryota* występuje dużo powtórzonych sekwencji zasad. Ponad 30% DNA stanowią sekwencje powtórzone co najmniej dwadzieścia razy. Sekwencje powtórzone w genomie człowieka można ogólnie podzielić na dwa typy: sekwencje rozproszone (*interspersed repetitive DNA*), gdzie powtórzone elementy są oddzielone od siebie innymi sekwencjami, oraz sekwencje powtórzone tandemowo (*tandemly repeated DNA*), w których fragmenty powtórzone ułożone są obok siebie i łączą się jak „głowa do ogona” np. A¹T²A³A⁴A⁵C⁶T⁷A¹T²A³A⁴A⁵C⁶T⁷... Cyfry oznaczają kolejność zasad azotowych.

Do rozproszonych sekwencji powtórzonych należą m.in. sekwencje SINE (*Short Interspersed Nuclear Elements*) – krótkie rozproszone elementy jądrowe o długości 100–500 par zasad, i LINE (*Long Interspersed Nuclear Elements*) – długie rozproszone elementy jądrowe wielkości kilku i więcej kb. Do tandemowych sekwencji powtórzonych zalicza się satelitarny DNA.

Do sekwencji SINE należą występujące u człowieka sekwencje Alu – nazwa pochodzi od endonukleazy restrykcyjnej AluI rozpoznającej określone miejsca w obrębie tej sekwencji. Są to fragmenty długości 280 par zasad, występujące w genomie człowieka wielokrotnie – około miliona razy. Sekwencje Alu należą do tzw. retrotranspozonów. Oznacza to, że transkrypty tych sekwencji ulegają odwrotnej transkrypcji i w postaci DNA integrują się z genomem w nowym miejscu. Również sekwencje LINE zaliczane są do retrotranspozonów.

Sekwencje ruchome (transpozony), należące do powtórzeń rozproszonych w genomie, to fragmenty DNA zdolne do przemieszczania się w obrębie genomu. Transpozycja DNA jest powszechna w ludzkim genomie, lecz najczęściej nie obejmuje sekwencji kodujących. Sekwencje ruchome zaliczane do rozproszonych sekwencji powtórzonych można podzielić na retropozony, transpozony i retrotranspozony.

Retropozony przemieszczają się w genomie jednej komórki w wyniku transkrypcji, syntezy DNA na matrycy RNA przy udziale odwrotnej transkryptazy, syntezy dwuniciowej formy komplementarnego DNA, integracji (DNA-RNA-DNA). Taki sposób przemieszczania nazywa się retropozycją, a sekwencje przemieszczane retropozonami.

Transpozony to sekwencje DNA ulegające bezpośredniej transpozycji fragmentu DNA w obrębie jednej komórki, z pominięciem RNA.

Retrotranspozony to fragmenty DNA, które mogą się przemieszczać nie tylko wewnątrz genomu danej komórki, ale również do genomu innych komórek. Transpozycja retrotranspozonów odbywa się podobnie jak retropozonów na drodze DNA-RNA-DNA. Retrotranspozony zawierają pełną informację genetyczną kodującą białka enzymatyczne niezbędne do retrotranspozycji wewnątrzkomórkowej, a nawet zewnątrzkomórkowej. Właściwości retrotranspozonu stają się zatem podobne do retrowirusów, stąd sekwencje te bywają często nazywane retrowirusopodobnymi.

W powtórzeniach tandemowych występują geny (rDNA) kodujące rybosomalny RNA (rRNA). Niemal u wszystkich *Eukaryota* geny te występują w więcej niż 100 kopiach (u człowieka ok. 450 kopii). W powtórzeniach (ok. 30–40 razy) występują również geny kodujące białka histonowe.

Do sekwencji tandemowych o wielkiej liczbie powtórzeń należy DNA satelitarny. Nazwa pochodzi od właściwości tego DNA w czasie ultrawierowania w gradiencie gęstości CsCl. Gęstość satelitarnego DNA znacznie się różni od gęstości głównego pasma DNA,

jest on bowiem cięższy lub lżejszy w zależności od stosunku par G+C do A+T. Ponieważ występują trzy frakcje satelitarne, ten rodzaj DNA podzielono na trzy typy: satelitarny I (lekki, z dużą zawartością A+T), satelitarny II i satelitarny III (bogaty w G+C).

DNA satelitarny różnych gatunków *Eukaryota* ma pewne cechy specyficzne, ale można stwierdzić również istnienie homologii międzygatunkowych. U człowieka wyróżniono 5 głównych grup satelitarnego DNA:

- **grupa 1** – to heterogenna grupa powtórzeń pięciu nukleotydów, stanowiąca główny składnik frakcji II i III po ultrawiwrowaniu. Występuje w heterochromatinie ramion długich chromosomu Y;
- **grupa 2** – bogata w A+T, to główny składnik I frakcji DNA, po wirowaniu składa się z powtórzeń sekwencji o długości 17 par zasad (typ A) lub 25 par zasad (typ B). Ten rodzaj DNA znajduje się w przycentromerowych regionach chromosomów oraz w ramionach długich chromosomu Y;
- **grupa 3** – bogata w G+C, składa się z powtórzeń o długości 70 lub 140 par zasad. Liczy ok. 4000 kopii;
- **grupa 4** – to satelitarny DNA występujący w regionach centromerowych wszystkich chromosomów. U człowieka stanowi ok. 3% genomu, a np. u małp ok. 30%;
- **grupa 5** – to DNA występujący w ramionach długich chromosomu Y, składa się z powtórzonych sekwencji o długości 2400 par zasad.

W genomie *Eukaryota* występują również małe bloki sekwencji satelitarnych o niewielkiej liczbie powtórzeń jednostki podstawowej, zwane minisatelitami. Pierwszą z tych sekwencji wykryto w intronie genu mioglobiny ludzkiej. Zawiera ona cztery powtórzenia sekwencji o długości 33 par zasad. Minisatelity zawierają powtórzenia tandemowe złożone z 9–80 pz. Duża zmienność minisatelitów wynika z różnej liczby powtórzeń danego motywu w określonym *locus*. Zmienność ta, występująca m.in. u człowieka, została określona jako polimorfizm liczb tandemowych

powtórzeń VNTR (*Variable Number Tandem Repeats*).

Analiza kilku takich *loci* danego osobnika technikami hybrydizacyjnymi pozwala uzyskać obraz dla niego charakterystyczny, określane jako „genetyczny odcisk palca” (*fingerprint*). Obraz ten jest unikatowy i może być wykorzystywany przy ustalaniu ojcostwa, identyfikacji sprawcy przestępstwa lub identyfikacji zwłok. Do podobnych celów można również wykorzystać analizę *loci* mikrosatelitarnych.

Mikrosatelity to powtórzenia tandemowe zawierające 10–50 powtórzeń o długości 1–6 par zasad. Są one wykorzystywane jako markery w diagnostyce wielu chorób dziedzicznych. Mikrosatelity są znacznie krótsze i częściej występują w genomie niż minisatelity. Wzrost zainteresowania tymi sekwencjami wynika z licznych doniesień o ich niestabilności w wielu typach nowotworów.

6.4

Genom mitochondrialny

Mitochondria to organelle wewnątrzkomórkowe dziedziczone prawie wyłącznie po matce. Udział mitochondrialnego DNA (mtDNA) plemnika w dziedziczeniu pozajądrowym wynosi zaledwie 0,1%, gdyż w komórce jajowej znajdują się dziesiątki tysięcy mitochondriów, natomiast w płemniku jest ich około dwudziestu.

Liczba mitochondriów w komórkach somatycznych człowieka jest różna w zależności od typu tkanki, średnio ok. 1000.

Najwięcej mitochondriów zawierają włókna mięśniowe poprzecznie prążkowane, mięsień sercowy, nerki oraz ośrodkowy układ nerwowy. Dlatego zespoły chorobowe dziedziczone mitochondrialnie to głównie schorzenia neurologiczne i encefalomiopatie.

Każde mitochondrium zawiera 2–10 cząsteczek DNA. DNA mitochondrialny nie ulega rekombinacji, niewielkie są również możliwości naprawy DNA w mitochondriach. Różnice w budowie genomu mitochondrialnego są więc wynikiem kumulujących się zmian mutacyjnych. Liczba mutacji w mitochondrialnym DNA jest ok. 10 razy większa niż w DNA jądrowym. Jest to głównie związane z tym, że w mitochondriach odbywa się proces tlenowego oddychania wewnątrzkomórkowego, podczas którego ok. 1–2% tlenu przemienia się w wolne rodniki tlenowe, zaliczane do czynników mutagennych.

Genom mitochondrialny człowieka jest zbudowany z dwuniciowej kolistej cząsteczki DNA zawierającej 16 569 par zasad, co stanowi mniej niż 1% całkowitego DNA w komórce. Jedną z nici tego genomu zawierającą większość genów określa się jako nić ciężka (H), drugą jako nić lekka (L). Genom mitochondrialny zawiera 22 geny kodujące tRNA, 2 geny rRNA oraz 13 regionów kodujących białka (cytochrom b, 3 podjednostki oksydazy cytochromowej, dwie podjednostki ATP-azy, 7 podjednostek dehydrogenazy NADH). DNA mitochondrialny człowieka nie zawiera intronów. Ekspresja większości genów w genomie mitochondrialnym odbywa się w tym samym kierunku, a geny tRNA leżą między genami kodującymi rRNA lub białka.

Prawie każda para zasad mitochondrialnego DNA wchodzi w skład genu zawierającego informację o syntezie białka czy też RNA. Z wyjątkiem pętli D-regionu, w którym następuje inicjacja replikacji DNA, nie więcej niż 87 z 16 569 par zasad może leżeć w regionie intercystronowym (między genami). W wielu przypadkach ostatnia zasada azotowa jednego genu przylega bezpośrednio do pierwszej zasady następnego genu, co świadczy o dużej oszczędności organizacji genomu mitochondrialnego. Niekiedy następuje tzw. nadpisanie (*overlap*) jednej zasady, dzięki czemu ostatnia zasada jednego genu jest również pierwszą zasadą następnego genu.

Mitochondria mają własny kod genetyczny. Okazało się, że pięć ramek odczytu nie ma kodonu terminacyjnego i kończą się one na U lub UA, a kodon *ochra* (UAA) jest tworzony dopiero przez poliadenylację transkrypty (mitochondrialny mRNA nabywa krótką sekwencję poli (A) na końcu 3'). Ponadto w trzech przypadkach kodonem terminacyjnym są sekwencje AGA lub AGG, które zwykle kodują argininę. Kodonami startowymi są natomiast sekwencje AUG, AUA czy też AUU.

Mutacje mitochondrialnego DNA są odpowiedzialne za wystąpienie wielu chorób, m.in. zespołu Kearnsa-Sayre'a – KSS (oftalmoplegia, zwyrodnienie barwnikowe siatkówki, kardiomiopatia), dziedzicznej neuropatii nerwu wzrokowego – LHON (zespół Lebera) oraz encefalomiopatii, kwasicy mleczanowej z objawami udaropodobnymi – MELAS. Mutacje w mtDNA mogą również odpowiadać za procesy starzenia się (rozdz. 33).

6.5

Aktualne osiągnięcia w dziedzinie poznania ludzkiego genomu

Historia odczytywania ludzkiego genomu sięga roku 1952, kiedy to James Watson opisuje DNA, z którego zbudowane są geny. Szersze prace nad określeniem kolejności zasad, z których zbudowane są nici DNA, przepadły na lata 80. i 90. XX w.

Od kilkunastu lat przodujące badania w zakresie poznania genomu człowieka prowadzi grupa naukowców z 18 państw związanych z Projektem Poznania Genomu Człowieka HGP (*Human Genome Project*).

Badaniami w ramach HGP kieruje Narodowe Centrum Badań Genomu Człowieka USA NCHGR (*National Center for Human Genome Research*) powstałe w 1988 r. przy Narodowym Instytucie Zdrowia NIH (*National Institutes of Health*). Projekt ten jest

bardzo kosztowny. Planowany budżet przewidywał 200 mln USD rocznie przez 15 lat.

Projekt Poznania Genomu Człowieka jest bez wątpienia jednym z największych przedsięwzięć badawczych ludzkości. Wielu naukowców z kilkunastu krajów podjęło wyzwanie poznania pełnego zapisu informacji zawartej w DNA, według której funkcjonuje cały organizm.

Najważniejsze kierunki realizacji HGP to:

- opracowanie map chromosomowych człowieka,
- utworzenie map i rozpoczęcie sekwencjonowania genów organizmów modelowych (*Escherichia coli*, drożdże *Saccharomyces cerevisiae*, nicienie *Caenorhabditis elegans*, muszka *Drosophila melanogaster* oraz mysz laboratoryjna), które mają ułatwić interpretację wyników uzyskanych podczas badania ludzkiego genomu,
- usprawnienie technik sekwencjonowania DNA,
- ułatwienie dostępu do nowych technologii,
- zbadanie implikacji etycznych, prawnych i społecznych HGP.

Naukowcy pracujący w ramach HGP oraz w firmie amerykańskiej Celera Genomics Corporation z Rockville doprowadzili do prawie całkowitego zsekwencjonowania genomu człowieka, posługując się odmiennymi metodami. Klasyczna technika sekwencjonowania stosowana w ramach HGP polegała na fragmentacji DNA chromosomów, określeniu kolejności nukleotydów w tych odcinkach i ustawieniu uzyskanych sekwencji DNA w odpowiedniej kolejności. Celera Genomics Corporation wprowadziła nowy sposób analizowania sekwencji DNA, nazwany metodą sekwencjonowania dubeltówkowego. Metoda ta polega na pocięciu DNA badanego organizmu na ogromną liczbę fragmentów, które są następnie sekwencjonowane bez zwracania uwagi na ich położenie względem siebie w chromosomach. Następnie dopasowuje się do siebie „teksty” poszczególnych fragmentów, tak by uzyskać kompletny zapis. Dopasowywanie sekwencji przeprowadza się komputerowo. Dzięki tej metodzie realiza-

cja celów badawczych HGP została znacznie przyspieszona.

Jak już wspomniano, w pracach nad poznaniem ludzkiego genomu uczestniczyły głównie dwie grupy badawcze: międzynarodowy zespół realizujący HGP oraz Celera Genomics Corporation. Naukowcom pracującym w ramach HGP udało się odczytać sekwencje ludzkiego genomu w 85% (do czerwca 2000 r.), natomiast Celera rozszyfrowała ponad 90% sekwencji genomu. Nie odczytano wówczas całego genomu człowieka. Już rok później firma prywatna deCODE udostępniła mapę genomu człowieka pięciokrotnie dokładniejszą. Firma ta podała również informacje o usunięciu 104 błędów powstałych w poprzednim mapowaniu. Pod koniec 1999 r. poznano sekwencję pierwszego chromosomu – chromosomu 22. W 2003 r. znano już sekwencje 7 chromosomów, w tym chromosomu Y. W marcu 2005 r. odczytano sekwencję chromosomu X, a w maju 2006 r. poznano sekwencję ostatniego z badanych chromosomów – chromosomu 1 (tab. 6.1). Poznanie sekwencji zapisu informacji genetycznej („liter” i „wyrazów”) nie oznacza jeszcze, że wiadomo, do jakiej funkcji organizmu one się odnoszą. Ustalenie sekwencji DNA jest tylko pierwszym krokiem do zbadania funkcjonowania komórek i organizmów. Kolejny etap prac badawczych nad genomem człowieka polega na rozróżnianiu genów, poznaniu ich funkcji i ustaleniu współdziałania genów. Nadal nie wiadomo, z ilu dokładnie genów składa się ludzki genom.

Do tej pory ustalono genomy bakterii *E. coli* i innych bakterii chorobotwórczych, m.in. *Helicobacter pylori* wywołującej chorobę wrzodową żołądka. Poznano również pełną sekwencję genomu nicienia *Caenorhabditis elegans*, muszki *Drosophila melanogaster*, kilku gatunków ryb, myszy, szczura i innych organizmów.

W najnowszych doniesieniach podano liczbę ok. 25 tys. genów (sekwencji kodujących białka) u człowieka. Jest to o 10 tys. mniej, niż ogłoszono w 2001 r., przedstawiając wyniki badań prowadzonych w ramach HGP.

Tabela 6.1 Wielkość chromosomów człowieka i całkowita liczba genów (wg Woźniak i Mila-Kierzenkowska na podstawie danych z *National Center for Biotechnology Information – NCBI*)

Chromosom	Całkowita liczba genów	Wielkość w Mb	Zidentyfikowane geny	Data poznania pełnej sekwencji
1	3186	247	ok. 90%	maj 2006
2	2093	243	ok. 95%	kwiecień 2005
3	1638	200	ok. 95%	kwiecień 2006
4	1300	191	ok. 95%	kwiecień 2005
5	1448	181	ok. 95%	wrzesień 2004
6	1843	171	ok. 95%	październik 2003
7	1722	159	ok. 95%	lipiec 2003
8	1162	146	ok. 95%	styczeń 2006
9	1394	140	ok. 85%	maj 2004
10	1259	135	ok. 95%	maj 2004
11	2000	134	ok. 95%	marzec 2006
12	1509	132	ok. 95%	marzec 2006
13	611	114	ok. 80%	marzec 2004
14	1420	106	ok. 80%	kwiecień 2003
15	1143	100	ok. 80%	marzec 2006
16	1270	89	ok. 85%	grudzień 2004
17	1650	79	ok. 95%	kwiecień 2006
18	480	76	ok. 95%	marzec 2004
19	1861	64	ok. 85%	marzec 2004
20	824	62	ok. 90%	luty 2001
21	386	47	ok. 70%	maj 2001
22	812	50	ok. 70%	grudzień 1999
X	1529	155	ok. 95%	marzec 2005
Y	344	56	ok. 50%	czerwiec 2003

Odkrycie sekwencji ludzkiego genomu budzi ogromne nadzieje. Szczegółowe poznanie „szyfru życia” zapisanego w genomie człowieka będzie można wykorzystać przy produkcji nowych leków i doskonaleniu metod leczenia, w tym terapii genowej. Osiągnięcia wyników badań HGP umożliwią zastosowanie skuteczniejszych niż dotąd metod diagnostycznych w wykrywaniu i zapobieganiu chorobom uwarunkowanym genetycznie.

Poznanie dokładnej mapy ludzkiego genomu, oprócz niezaprzeczalnych korzyści poznawczych i czysto medycznych, niesie ze sobą również wiele dylematów natury bioetycznej i prawnej. Jednym z wielu problemów jest pytanie, kto powinien mieć dostęp do informacji o stanie materiału genetycznego danej osoby, ponieważ niekontrolowane przedostanie się takiej informacji, np. do pracodawcy czy też firm ubezpieczeniowych, może mieć dla wielu osób negatywne konsekwencje.

Problemom tym mogą zaradzić tylko odpowiednio przeprowadzone regulacje prawne. Zagadnienia dotyczące kwestii prawno-etycznych, związanych z badaniami genetycznymi u człowieka, są treścią licznych artykułów prawnych zawartych w konwencjach przyjętych przez Komitet Ministrów Rady Europy, takich jak:

- Konwencja o Ochronie Praw Człowieka i Godności Istoty Ludzkiej wobec Zastosowań Biologii i Medycyny (Konwencja o Prawach Człowieka i Biomedycynie z dnia 19 listopada 1996 r.),
- Powszechna Deklaracja o Genomie Ludzkim i Prawach Człowieka przyjęta na 29 sesji Konferencji Generalnej UNESCO w dniu 11 listopada 1997 r.

Informacje dotyczące postępu badań nad poznaniem ludzkiego genomu dostępne są w internecie. Można je znaleźć np. na stronie funkcjonującego jako agencja rządowa National Human Genome Research Institute w USA (<http://www.genome.gov/>). Aktualne dane dotyczące chromosomalnej lokalizacji genów, diagnostyki chorób, stanu zaawansowania mapy genów oraz budowy genomów innych organizmów dostępne są m.in. na stronie National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/>.

6.6

Mapowanie genomu

Mapowanie genomu polega na przypisaniu konkretnym genom określonych pozycji w poszczególnych chromosomach. W celu scharakteryzowania genomu ludzkiego stosuje się trzy rodzaje map:

- genetyczne (mapy sprzężeń),
- cytogenetyczne,
- fizyczne.

Mapowanie genetyczne opiera się na pomiarze tendencji dwóch nieallelicznych genów do wspólnego segregowania w czasie mejozy. Mapy sprzężeń pozwalają na ustalenie względnego położenia genów na chro-

mosomach. Mapy cytogenetyczne odwziewiedlają pasmowy układ prążków charakterystyczny dla danego chromosomu w zależności od zastosowanej metody wybarwienia. Położenie genu w specyficznym miejscu (*locus*) na chromosomie ukazują natomiast mapy fizyczne, które wymagają bezpośredniego badania cząsteczek DNA i mogą mieć różny stopień rozdzielczości. Wszystkie rodzaje map uzupełniają się i tworzą podstawę do dalszych badań w celu otrzymania kompletnej sekwencji ludzkiego DNA. Poznanie pełnej sekwencji DNA poszczególnych chromosomów stanowi najwyższy stopień rozszyfrowania genomu ludzkiego (ryc. 6.1).

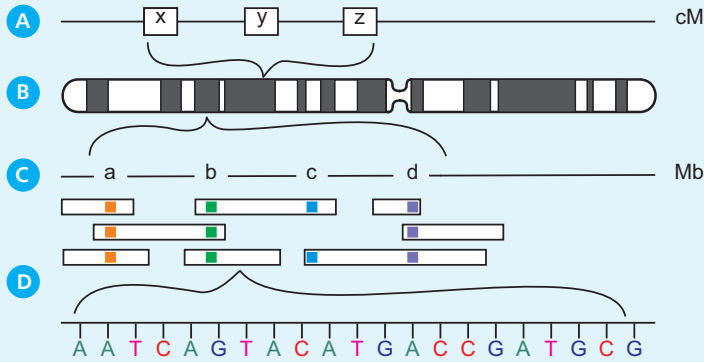
6.6.1

Mapy genetyczne (mapy sprzężeń)

Podstawą mapowania genetycznego jest zjawisko rekombinacji odkryte przez Thomasa Morgana na początku poprzedniego wieku. Mapy genetyczne sporządza się, analizując sprzężenia na podstawie oceny częstości rekombinacji między genami leżącymi na tym samym chromosomie.

Dwa geny są sprzężone ze sobą, jeżeli ich *loci* leżą na tym samym chromosomie. Allele tych dwóch genów mają tendencję do przechodzenia razem do tej samej gamety. Jeżeli lokalizacja jednego genu jest znana, to drugi gen łatwo zlokalizować w danym obszarze chromosomu. W analizie sprzężeń brane są pod uwagę dwa miejsca (*loci*) na chromosomie. W jednym z nich zlokalizowany jest gen określający badaną cechę, w drugim zaś obecny jest gen markerowy.

Markery genetyczne to charakterystyczne sekwencje, które są silnie sprzężone z badanym genem i dziedziczone są razem z nim, a rekombinacja między nimi zachodzi niezmiernie rzadko. W analizie sprzężeń genów można wykorzystać wiele różnych markerów, muszą one jednak spełniać podstawowy warunek. Podobnie jak badane geny, markery muszą występować w przynajmniej dwóch alternatywnych odmianach fenotypowych,



Ryc. 6.1 Różne stopnie poznania genomu. **A.** Mapa genetyczna – odległości między genami (x, y, z) oparte są na częstości *crossing-over* i podane w centymorganach (cM); **B.** Mapa cytogenetyczna – ukazuje charakterystyczny układ prążków, w których zlokalizowane są poszczególne geny; **C.** Mapa fizyczna – położenie specyficznych sekwencji (a-d) wyznaczone jest np. przez kolejność charakterystycznych miejsc restrykcji i podane w milionach par zasad (Mb); **D.** – pełna sekwencja zasad określonego odcinka DNA.

przy czym każdy fenotyp określany jest przez inny allel odpowiadającego mu genu. Najczęściej wykorzystywanymi obecnie typami markerów genetycznych są:

- polimorfizmy długości odcinków restrykcyjnych – RFLP (*Restriction Fragment Length Polymorphism*),
- polimorfizmy długości prostych sekwencji – SLP (*Simple Sequence Length Polymorphism*),
- polimorfizmy punktowe – SNP (*Single Nucleotide Polymorphism*).

Jeżeli dwa geny są całkowicie sprzężone ze sobą, to podczas mejozy przechodzą do gamety zawsze razem. W rzeczywistości jednak niewiele genów wykazuje pełne sprzężenie. Jeżeli czasem para genów dziedziczy się łącznie, a czasem w sposób przypominający niezależne dziedziczenie cech, to mówi się o sprzężeniu częściowym. Wyjaśnieniem zjawiska sprzężenia częściowego jest proces rekombinacji homologicznej (*crossing-over*), który zachodzi w mejozie podczas tworzenia się gamet. W spermatogenezie u człowieka *crossing-over* występuje przeciętnie około 52 razy (od 1 do 6 na chromosom w zależności od jego długości), a w czasie oogenezy zachodzi nawet dwa razy

częściej (rozd. 2). Analizując fenotyp potomstwa, można określić, czy nastąpiła rekombinacja czy też nie (tab. 6.2).

Zakładając, że proces *crossing-over* zachodzi w losowo wybranych miejscach chromosomu, to częstość rekombinacji między dwoma genami będzie odbiciem odległości między tymi dwoma *loci* genów. Dwa blisko położone od siebie geny są rozdzielane przez *crossing-over* rzadziej niż dwa geny bardziej od siebie oddalone. Częstość, z jaką dwa geny zostają rekombinowane, jest wprost proporcjonalna do ich odległości na chromosomie, a zatem jest miarą odległości między tymi genami. W wyniku analizy częstości rekombinacji dla każdej grupy genów sprzężonych można ustalić ich rozmieszczenie wzdłuż chromosomów.

Przykład

Częstość zachodzenia *crossing-over* między trójką genów (X, Y, Z) występujących na jednym chromosomie jest następująca: X – Y: 12%, Y – Z: 8% i X – Z: 4%. Jaka jest kolejność wymienionych genów?

